CERTIFICATE OF EFS FILING UNDER 37 CFR §1.8

I hereby certify that this correspondence is being filed electronically with the U.S. Patent and Trademark Office, via the EFS pursuant to 37 CFR §1.8, on the below date:

Date: October 14, 2008 Name: James P. Naughton Signature: J. Naughton

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:

Shingo Kiuchi, et al.

Serial No. 10/730,767

Examiner: James S. Wozniak

Filing Date: December 8, 2003

Group Art Unit No.: 2626

For SPEECH RECOGNITION PERFORMANCE IMPROVEMENT METHOD AND SPEECH RECOGNITION DEVICE

Confirmation No.: 3437

Attorney Docket No. 9333/361

APPEAL BRIEF

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

This is an appeal from the Final Office Action dated May 27, 2008. Appellants respectfully appeal the final rejection entered by the Examiner and provide this Appeal Brief in support thereof.

I. REAL PARTY IN INTEREST

It is believed that Alpine Electronics, Inc. is the real party in interest in this Appeal pursuant to the assignment of the above-identified application to Alpine Electronics, Inc.

II. RELATED APPEALS AND INTERFERENCES

The undersigned, James P. Naughton, is not aware of any other prior or pending appeals, interferences, or judicial proceedings which may be related to, directly affect, or be directly affected by or have a bearing on the Board's decision in the pending Appeal.

III. STATUS OF CLAIMS

The status of the claims is as follows:

1) Claims 1, 3-8, 12-15, 17, 18 and 20 are present and active in the application; claims 2, 9-11, 16 and 19 are cancelled.

2) Claims 1, 3-8, 12-15, 17, 18 and 20 are finally rejected under 35 U.S.C. § 103(a) as unpatentable over U.S. Patent No. 4,885,791 ("Fujii") in view of U.S. Patent No. 6,975,993 ("Kieller") and further in view of U.S. Patent No. 6,324,509 ("Bi").

3) The rejections of claims 1, 3-8, 12-15, 17, 18 and 20 are being appealed.

IV. STATUS OF AMENDMENTS

An Amendment was filed on July 23, 2008 in response to the Final Office Action dated May 27, 2008. Pursuant to an Advisory Action dated August 14, 2008, the Amendment was entered and considered by the examiner but was not found to place the application in condition for allowance.

The claims in the attached Claims Appendix include the claim amendments made in the Amendment filed on July 23, 2008.

V. SUMMARY OF CLAIMED SUBJECT MATTER

An understanding of the invention of independent claims 1, 8 and 15, and their respective dependent claims can be obtained upon a review of the embodiments of the invention described below and illustrated in the figures of the specification.

Appellants' invention relates to a device and method for improving speech recognition performance in a noisy environment. When noise is superimposed on speech data that is entered to a speech recognition system, the recognized result may change. (See, e.g., p. 1, lines 5-9; p. 1, line 25 to p. 2, line 31.)

One object of the invention is to improve speech recognition performance without changing the speech recognition engine itself. In the speech recognition device of one embodiment, a plurality of pieces of speech data whose start positions of preceding non-speech regions differ are generated from speech data for which speech recognition is to be performed. Speech recognition is performed by using each of the pieces of speech data, and the most numerous recognized result from among a plurality of obtained recognized results is provided as an output. As a result, since the start position of the non-speech region is shifted, although there may happen to be speech data which is recognized incorrectly, if a large number of pieces of speech data are recognized and the numbers compared, the number of cases in which the speech data is recognized correctly becomes the most numerous. Therefore, by providing the result which is recognized most often, the recognition performance can be improved without changing the recognition engine. (E.g., p. 3, lines 6-19.)

In order to generate a plurality of pieces of speech data whose start positions of non-speech regions differ, the start position of the non-speech region is shifted in sequence from the start position of the speech region to a preceding position by a predetermined time. That is, the input speech signal is A/D-converted at a predetermined sampling speed, and this speech signal may be stored in a buffer in the order of sampling. Then, a plurality of pieces of speech data whose start positions of non-speech regions differ is generated by changing the position at which reading from the speech buffer starts. (E.g., p. 3, lines 20-27.)

The speech recognition process of each of the pieces of speech data may be performed by one speech recognition engine, but this approach takes time. In order to shorten the processing time, in an alternative embodiment a speech recognition engine is provided so as to correspond to each of the pieces of speech data whose start positions of the non-speech regions differ, and the most numerous recognized result from among the recognized results of each speech recognition engine is supplied as an output. By providing

a speech recognition engine in such a manner as to correspond to each of the plurality of pieces of speech data, a speech recognition result can be obtained at a high speed, and moreover, recognition performance can be improved. (E.g., p. 3, lines 28 to p. 4, line 3; p. 4, lines 13-16.)

There are no means-plus-function terms or step-plus-function terms recited in the claims on appeal.

Below are copies of the independent claims. After each limitation is a citation to the specification and drawings which apply to the limitation.

Claim 1

1. A method for use with a speech recognition device for improving speech recognition performance, said method comprising: (e.g., p. 1, lines 5-9; p. 3, line 6 to p. 4, line 22)

identifying a start position of a speech region of speech data for which speech recognition is to be performed; (e.g., p. 5, line 20 to p. 6, line 4; p. 6, lines 22-27; Figs. 1-2, operation of the speech recognition engine 17)

generating, from said speech data for which speech recognition is to be performed, a plurality of pieces of speech data including said speech region and a varying period of a preceding non-speech region, where start positions of non-speech regions differ for the plurality of pieces of speech data; (e.g., p. 5, line 6 to p. 6, line 21; Figs. 1-3)

performing speech recognition using each of said pieces of speech data to obtain a plurality of recognized results; and (e.g., p. 6, line 22 to p. 7, line 2)

identifying a most numerous recognized result from among the plurality of obtained recognized results; (e.g., p. 7, lines 3-11; Fig. 4, operation of the totaling/comparison section 19 from Fig. 1)

wherein, by sequentially shifting the start position of said non-speech region from the start position of the speech region back to a position preceding by a predetermined time, a plurality of pieces of speech data whose start positions of non-speech regions differ

are generated from said speech data for which speech recognition is to be performed. (e.g., p. 6, lines 5-21; Figs. 2-3)

Claim 8

8. A speech recognition device comprising: (e.g., p. 1, lines 5-9; p. 3, line 6 to p. 4, line 22)

a speech data generation section for identifying a start position of a speech region of speech data for which speech recognition is to be performed and generating, from said speech data for which speech recognition is to be performed, a plurality of pieces of speech data including said speech region and a varying period of a preceding non-speech region, where start positions of non-speech regions differ for the plurality of pieces of speech data; (e.g., p. 5, line 6 to p. 6, line 27; Figs. 1-3, operation of the speech recognition engine 17)

a speech recognition engine for performing speech recognition on each of said pieces of speech data to obtain a plurality of recognized results; and (e.g., p. 6, line 22 to p. 7, line 2)

a recognized result selection section for selecting a most numerous recognized result from among the plurality of obtained recognized results; (e.g., p. 7, lines 3-11; Fig. 4, operation of the totaling/comparison section 19 from Fig. 1)

wherein said speech data generation section generates a plurality of pieces of speech data whose start positions of non-speech regions differ from speech data for which speech recognition is to be performed by sequentially shifting the start position of said non-speech region to a position preceding by a predetermined time from the start position of the speech region. (e.g., p. 6, lines 5-21; Figs. 2-3)

Claim 15

15. A speech recognition device for improving speech recognition performance, said speech recognition device comprising: (e.g., p. 1, lines 5-9; p. 3, line 6 to p. 4, line 22; p. 7, lines 22-26; p. 8, lines 7-9; Fig. 5)

a speech data generation section for identifying a start position of a speech region of speech data for which speech recognition is to be performed and generating, from said speech data for which speech recognition is to be performed, a plurality of pieces of speech data including said speech region and a varying period of a preceding non-speech region, where start positions of non-speech regions differ for the plurality of pieces of speech data; (e.g., p. 5, line 6 to p. 6, line 27; p. 7, lines 22-26; Figs. 1-3, 5; operation of the speech recognition engine 17)

a speech recognition engine, for performing speech recognition on the speech data, provided for each of a plurality of pieces of speech data whose start positions of non-speech regions differ in order to obtain a plurality of recognized results; and (e.g., p. 6, line 22 to p. 7, line 2; p.7, line 29 to p. 8, line 3; Fig. 5)

a recognized result section for selecting and providing as an output a most numerous recognized result from among the plurality of obtained recognized results; (e.g., p. 7, lines 3-11; p. 8, lines 4-6; Figs. 4-5, operation of the totaling/comparison section 19 from Fig. 1)

wherein said speech data generation section generates, from speech data for which speech recognition is to be performed, a plurality of pieces of speech data whose start positions of non-speech regions differ, by sequentially shifting the start position of said non-speech region from the start position of the speech region back to a position preceding by a predetermined time. (e.g., p. 6, lines 5-21; p. 7, lines 22-26; Figs. 2-3, 5)

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

The grounds of rejection presented for review are the rejections of claims 1, 3-8, 12-15, 17, 18 and 20 under 35 U.S.C. § 103(a) as being obvious over Fujii, Keiller and Bi.

VII. ARGUMENT

Reversal of the rejections of claims 1, 3-8, 12-15, 17, 18 and 20 is respectfully requested.

A. Independent Claims 1, 8 and 15

Each of the independent claims is rejected under 35 U.S.C. § 103(a) as obvious over Fujii, Kieller and Bi, with Fujii being the primary reference.

Method claim 1 recites a method for improving speech recognition performance in a speech recognition device. First, a start position of a speech region of speech data for which speech recognition is to be performed is identified. Then, a plurality of pieces of speech data including the speech region and a varying period of a preceding non-speech region is generated, where the start positions of the non-speech regions differ for the plurality of pieces of speech data.

Speech recognition is then performed using each of the pieces of speech data to obtain a plurality of recognized results, and the most numerous recognized result from among the plurality of obtained recognized results is identified as the speech recognition result. In this process, the plurality of pieces of speech data whose start positions of non-speech regions differ are generated from the speech data by sequentially shifting the start position of the non-speech region from the start position of the speech region back to a position preceding by a predetermined time.

Independent claim 8 is directed to a speech recognition device and recites limitations corresponding to the limitations in claim 1. Independent device claim 15 also recites corresponding limitations, except that a speech recognition engine is provided for each of the plurality of pieces of speech data on which speech recognition is to be performed.

Appellants submit that the cited art does not disclose or suggest the claimed invention. Fujii describes a speech recognition apparatus that attempts to reduce erroneous recognition due to noise by detecting one or more different speech periods (e.g., at Abstract; col. 4, lines 7-16). The Office Action acknowledges that Fujii does not generate a recognition result based on the most frequent recognized result, but instead selects a best overall candidate word based on pattern matching. Keiller describes a speech recognition system in which a single piece of speech data is passed through a plurality of speech recognition engines, and a most commonly occurring recognition result is selected as the most likely interpretation (e.g., col. 2, lines 4-8; col. 20, line 47 to col. 21, line 10), but

Keiller does not describe generating a plurality of pieces of speech data with different start positions.

The Office Action asserts that it would have been obvious to combine these features of the prior art to result in Appellants' independent claims, although the prior art references do not suggest this combination and the Office Action does not articulate any specific rationale to make this combination. This is improper. Factual findings by the Examiner and articulated reasoning are necessary underpinnings to establish obviousness and must be made explicit. *KSR Int'l Co. v. Teleflex Inc.*, 127 S. Ct. 1727, 1741 (2007); Examination Guidelines for Determining Obviousness, 72 Fed. Reg. 57,526, 57,528 (Oct. 10, 2007); 35 U.S.C. § 132. A mere conclusory statement cannot support the legal conclusion of obviousness. 127 S. Ct. at 1741. Rather, the Examiner must identify how a person of ordinary skill in the art would, by known methods, combine the elements in the way the claimed invention does. *Id.*; 72 Fed. Reg. 57,526, 57,528 (Oct. 10, 2007). A balanced review of the cited references show that they cannot reasonably be combined in the way proposed by the Office Action.

The primary reference, Fujii, discloses that errors in the detection of the speech period can occur due to noise, and Fujii addresses this problem not by attempting to determine the actual starting point of speech, but rather by applying the recognition algorithm to multiple "proposed speech periods" (e.g., col. 2, lines 16-18, 38-56; col. 14, lines 7-16; col. 8, lines 24-31). This way, Fujii does not require the detection of the actual speech period (col. 3, lines 21-22). Fujii does not know and does not need to know the actual starting point of speech. In contrast, Appellants' invention first identifies a start position of the speech region and then adds varying periods of the preceding non-speech region. Moreover, whereas Fujii is concerned with the problem of detecting the actual speech period boundaries in the presence of noise, Appellants' invention is concerned with how a varying level of noise may affect recognition accuracy (even if the speech period is known) (Application, e.g., at p. 1, line 25 to p. 2, line 27; p. 4, lines 17-22).

As the Office Action concedes (at pg. 6), Fujii does not disclose obtaining different speech period segments by shifting backwards. Indeed, Fujii is silent on how specifically to determine different beginning and terminating end points of the proposed speech periods.

Certainly, Fujii does not disclose including varying periods of a preceding non-speech region as in Appellants' claimed invention.

The Office Action cites Bi as disclosing the allegedly "well known means of achieving the multiple speech data periods" by sequentially shifting back from a determined starting point by a predetermined time. Appellants respectfully disagree. Bi, entitled "Method and Apparatus for Accurate Endpointing of Speech in the Presence of Noise," describes a detailed process that ends by determining a single set of start and stop points of a speech period. The passage in Bi relied upon by the Office Action (col. 5, lines 13-30) is not applicable. This passage merely notes that the signal data are stored in a buffer because the processor performs real-time processing but must be able to look back a certain number of speech frames.

In particular, Bi does not describe sequentially shifting backwards to obtain a plurality of starting points. Bi uses a first signal-to-noise ratio (SNR) threshold value to identify a "first starting point" and a "first ending point" of calculation that are not endpoints of the speech data but are instead only interim calculation points ("PRE_START" and "PRE_END"). Then, Bi uses a second, smaller SNR threshold value to determine the "actual" starting and ending points of the speech data (e.g., Abstract, col. 4, lines 37-57; col. 6, lines 15-39; col. 7, lines 24-30).

Bi notes that in conventional voice recognition devices, the endpoint detector relies upon a single SNR threshold to determine the endpoints of a piece of speech. However, setting the SNR threshold too low may make the device too sensitive to background noise, whereas setting the SNR threshold too high may miss part of the beginning or ending of speech (col. 2, lines 21-37). Bi, in contrast, "uses multiple, adaptive SNR thresholds to accurately detect the endpoints of speech in the presence of background noise" (col. 2, lines 42-44). Thus, Bi explicitly states that it only determines one actual starting point and one actual ending point.

Bi does not "consider multiple starting points" or disclose "different possible starting points" as the Office Action and the Advisory Action assert. Further, the beginning point of the Bi's process is not the start of the speech period (as in Applicants' invention), but rather is simply a point where the SNR reaches a first arbitrary threshold value. The interim

calculation points in Bi's process are only that - -interim. They are not "proposed" or "likely" starting points as asserted in the Advisory Action. In fact, the interim calculation points in Bi are not used as starting points; they are merely intermediate calculations on the way to determining the single set of endpoints of speech data which can then be used in a speech recognition process.

Also, Fujii and Bi cannot be combined as suggested by the Office Action and the Advisory Action because the references are incompatible. Bi's process is directed at finding the speech period, but a key feature of Fujii is not needing to determine or know the actual speech period. If the actual speech period were known from Bi's process, Fujii's process would not be needed.

In short, none of the cited references describes or suggests determining multiple different start positions of a plurality of pieces of speech data by sequentially shifting back from a start position of the speech region of speech data by a predetermined time. Thus, even if the references were to be combined, at least this feature of Appellants' invention would still be missing. Any other conclusion would be based on hindsight analysis in view of Appellants' claimed invention, which is to be avoided. "A factfinder should be aware, of course, of the distortion caused by hindsight bias and must be cautious of arguments reliant upon *ex post* reasoning." *KSR Int'l. Co. v. Teleflex Inc.*, 127 S. Ct. 1727, 1742 (2007).

B. Dependent Claim 3

Dependent claim 3 adds that the information of the start position of the speech region is provided by a speech recognition engine which performs the speech recognition. The Office Action asserts that the endpoint detector 22 of Bi discloses this feature. However, Bi's endpoint detector determines the single set of endpoints after conducting its iterative process, whereas Appellants' claimed start position of the speech region is identified first and then used to generate the plurality of pieces of speech data.

C. Dependent Claims 4, 12 and 20

These dependent claims add a feature that the information of the start position of the speech region is obtained by performing a recognition process on a first speech data by using the speech recognition engine, or is obtained by averaging speech data for several pieces of

data from the start which have been subjected to the recognition processing. Again, the Office Action relies on Bi to provide this feature, as for claim 3. Appellants disagree with this rejection as for claim 3. The rejections highlight a key distinction of the claimed invention over the cited art. Bi determines a single set of speech endpoints only after conducting the iterative process relied on by the Office Action. Appellants' claimed invention, to the contrary, begins with the start position of the speech region and proceeds from there. The Office Action can't have it both ways.

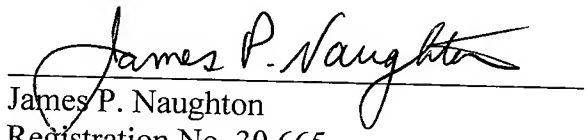
D. Dependent claims 5, 13 and 17

These dependent claims add a feature that a plurality of pieces of speech data whose start positions of non-speech regions differ are generated in such a manner that analog-to-digital conversion is performed on the input signal at a predetermined sampling time interval, the speech signal is stored in sequence in a speech buffer in an order of sampling, and a position at which reading from the speech buffer starts is changed. Contrary to the conclusory assertion of the Office Action, the cited art does not in any way disclose that the start positions of non-speech regions of a plurality of pieces of speech data are determined by changed the reading position in a speech buffer. The cited prior art is silent on this feature.

VIII. CONCLUSION

The cited references do not provide a valid basis for an obviousness rejection of the present claims. Therefore, Appellants submit that the claimed invention is patentable over the cited references, the Examiner's rejection should be REVERSED, and the claims should be ALLOWED.

Respectfully submitted,



James P. Naughton
Registration No. 30,665
Attorney for Appellants

BRINKS HOFER GILSON & LIONE
P.O. BOX 10395
CHICAGO, ILLINOIS 60610
(312) 321-4200
Dated: October 14, 2008

CLAIMS APPENDIX

1. A method for use with a speech recognition device for improving speech recognition performance, said method comprising:

identifying a start position of a speech region of speech data for which speech recognition is to be performed;

generating, from said speech data for which speech recognition is to be performed, a plurality of pieces of speech data including said speech region and a varying period of a preceding non-speech region, where start positions of non-speech regions differ for the plurality of pieces of speech data;

performing speech recognition using each of said pieces of speech data to obtain a plurality of recognized results; and

identifying a most numerous recognized result from among the plurality of obtained recognized results;

wherein, by sequentially shifting the start position of said non-speech region from the start position of the speech region back to a position preceding by a predetermined time, a plurality of pieces of speech data whose start positions of non-speech regions differ are generated from said speech data for which speech recognition is to be performed.

2. (Cancelled)

3. A method according to Claim 1, wherein the information of the start position of said speech region is provided by a speech recognition engine which performs said speech recognition.

4. A method according to Claim 3, wherein the information of the start position of said speech region is obtained by performing a recognition process on a first speech data by using said speech recognition engine, or is obtained by averaging speech data for several pieces of data from the start which have been subjected to the recognition processing.

5. A method according to Claim 1, wherein a plurality of pieces of speech data whose start positions of non-speech regions differ are generated in such a manner that analog-to-digital conversion is performed on the input signal at a predetermined sampling time interval, the speech signal is stored in sequence in a speech buffer in an order of sampling, and a position at which reading from the speech buffer starts is changed.

6. A method according to Claim 5, wherein said predetermined sampling time interval is for one piece of sampling data.

7. A method according to Claim 1, wherein a speech recognition engine is provided for each of a plurality of pieces of speech data whose start positions of non-speech regions differ, and the most numerous recognized result from among the recognized results of each speech recognition engine is identified.

8. A speech recognition device comprising:

a speech data generation section for identifying a start position of a speech region of speech data for which speech recognition is to be performed and generating, from said speech data for which speech recognition is to be performed, a plurality of pieces of speech data including said speech region and a varying period of a preceding non-speech region, where start positions of non-speech regions differ for the plurality of pieces of speech data;

a speech recognition engine for performing speech recognition on each of said pieces of speech data to obtain a plurality of recognized results; and

a recognized result selection section for selecting a most numerous recognized result from among the plurality of obtained recognized results;

wherein said speech data generation section generates a plurality of pieces of speech data whose start positions of non-speech regions differ from speech data for which speech recognition is to be performed by sequentially shifting the start position of said non-speech region to a position preceding by a predetermined time from the start position of the speech region.

9-11. (Cancelled)

12. A speech recognition device according to Claim 8, wherein the information of said start position of the speech region is obtained by performing a recognition process on a first speech data by using said speech recognition engine, or is obtained by averaging data of speech data for several pieces of data from the start, which have been subjected to the recognition processing.

13. A speech recognition device according to Claim 8, further comprising:

an analog to digital converter for converting an input speech signal from analog to digital at a predetermined sampling time interval; and

a speech buffer for storing the converted speech data in an order of sampling,

wherein said speech data generation section generates a plurality of pieces of speech data whose start positions of non-speech regions differ, by changing positions at which reading from the speech buffer starts.

14. A speech recognition device according to Claim 13, wherein said predetermined sampling time interval is for one piece of sampling data.

15. A speech recognition device for improving speech recognition performance, said speech recognition device comprising:

a speech data generation section for identifying a start position of a speech region of speech data for which speech recognition is to be performed and generating, from said speech data for which speech recognition is to be performed, a plurality of pieces of speech data including said speech region and a varying period of a preceding non-speech region, where start positions of non-speech regions differ for the plurality of pieces of speech data;

a speech recognition engine, for performing speech recognition on the speech data, provided for each of a plurality of pieces of speech data whose start positions of non-speech regions differ in order to obtain a plurality of recognized results; and

a recognized result section for selecting and providing as an output a most numerous recognized result from among the plurality of obtained recognized results;

wherein said speech data generation section generates, from speech data for which speech recognition is to be performed, a plurality of pieces of speech data whose start positions of non-speech regions differ, by sequentially shifting the start position of said non-speech region from the start position of the speech region back to a position preceding by a predetermined time.

16. (Cancelled)

17. A speech recognition device according to Claim 15, further comprising:

an analog to digital converter for converting an input speech signal from analog to digital at a predetermined sampling time interval; and

a speech buffer for storing the converted speech data in an order of sampling,

wherein said speech data generation section generates a plurality of pieces of speech data whose start positions of non-speech regions differ, by changing a reading position from the speech buffer, and provides the speech data to each speech recognition engine.

18. A speech recognition device according to Claim 17, wherein said predetermined sampling time interval is for one piece of sampling data.

19. (Cancelled)

20. A speech recognition device according to Claim 15, wherein the information of the start position of said speech region is obtained by performing a recognition process on a first speech data by using said speech recognition engine, or is obtained by averaging data of speech data for several pieces of data from the start, which have been subjected to the recognition processing.

EVIDENCE APPENDIX

None.

RELATED PROCEEDINGS APPENDIX

None.